

Prediksi *Downlink Throughput* Menggunakan Metode *Random Forest Regression* Pada Jaringan 4G LTE Berdasarkan Data *Drive Test*.

Riyadh Muhammad Adhim^{1*}, Vera Noviana Sulistyawan¹, Sri Sukamta¹

¹IJurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang, Semarang, Indonesia

Keywords:

Downlink Throughput;
4G LTE;
Random Forest Regression;
Drive Test;
Machine Learning.

Correspondent Email:

riyadhmuhammadadhim2002@gmail.com

Abstrak. Meningkatnya kompleksitas jaringan 4G LTE menuntut manajemen *Quality of Service* (QoS) yang proaktif, di mana kemampuan prediksi *downlink throughput* menjadi krusial bagi operator untuk menjamin pengalaman pengguna. Penelitian ini bertujuan membangun dan mengoptimalkan model prediksi *downlink throughput* menggunakan metode *Random Forest Regression* berdasarkan data primer hasil *drive test*. Data yang digunakan mencakup parameter kualitas sinyal (*Reference Signal Received Power*, *Reference Signal Received Quality*, *Signal-to-Interference-plus-Noise Ratio*) dan data lokasi sebagai fitur input, yang diolah melalui perbandingan metode pra-pemrosesan, optimisasi hiperparameter, dan analisis rasio pembagian data untuk menemukan konfigurasi model paling akurat. Kinerja model dievaluasi menggunakan metrik *Root Mean Square Error* (RMSE). Hasil penelitian menunjukkan bahwa *Random Forest Regression* berhasil membangun model prediksi yang efektif, di mana konfigurasi terbaik dicapai dengan menggunakan data seluruh fitur yang tersedia, rasio pembagian data 70:30, dan jumlah *trees* sebanyak 700. Model optimal ini mampu mencapai tingkat kesalahan (RMSE) sebesar 9201.575 kbps. Lebih lanjut, uji stabilitas melalui 1000 iterasi mengonfirmasi keandalan model, dengan selisih rata-rata antara nilai aktual (32222.05 kbps) dan prediksi (32253.18 kbps) yang sangat kecil, yaitu hanya 31.13 kbps. Temuan ini membuktikan bahwa model yang dikembangkan akurat, stabil, dan dapat diandalkan untuk estimasi *throughput* jaringan secara konsisten.



Copyright © [JPI](http://www.jpi.uns.ac.id) (Jurnal Profesi Insinyur Universitas Lampung).

Abstract. The increasing complexity of 4G LTE networks demands proactive *Quality of Service* (QoS) management, where the ability to predict *downlink throughput* is crucial for operators to guarantee user experience. This research aims to build and optimize a *downlink throughput* prediction model using the *Random Forest Regression* method based on primary data from *drive tests*. The data used includes signal quality parameters (*Reference Signal Received Power*, *Reference Signal Received Quality*, *Signal-to-Interference-plus-Noise Ratio*) and location data as input features. This data was processed by comparing pre-processing methods, hyperparameter optimization, and data split ratio analysis to find the most accurate model configuration. Model performance was evaluated using the *Root Mean Square Error* (RMSE) metric. The results show that *Random Forest Regression* successfully built an effective prediction model. The best configuration was achieved using all available features, a 70:30 data split ratio, and 700 trees. This optimal model achieved an error rate (RMSE) of 9201.575 kbps. Furthermore, a stability test through 1000 iterations confirmed the model's reliability, with a very small average difference of only 31.13 kbps between the actual (32222.05 kbps) and predicted (32253.18 kbps) values. These

findings prove that the developed model is accurate, stable, and reliable for consistent network throughput estimation.

1. PENDAHULUAN

Di tengah era transformasi digital yang pesat, jaringan seluler menghadapi peningkatan kompleksitas yang signifikan. Kualitas Layanan (QoS) menjadi tolok ukur krusial yang merefleksikan performa jaringan dari perspektif pengguna, terutama untuk layanan data seperti *streaming* dan komunikasi digital lainnya. Berbagai penelitian di Indonesia berfokus pada evaluasi dan pengukuran QoS untuk memetakan kualitas layanan di berbagai wilayah [1]. Studi oleh Simanjuntak dkk. di Karawang melakukan pengukuran parameter kunci seperti *throughput*, *delay*, dan *jitter* untuk membandingkan performa antar penyedia layanan seluler [2]. Kajian-kajian semacam ini menegaskan adanya variasi kualitas jaringan yang signifikan di lapangan, sehingga memperkuat kebutuhan akan pendekatan manajemen yang lebih inovatif dan proaktif untuk menjaga kepuasan pengguna.

Menghadapi tantangan ini, pendekatan tradisional dalam manajemen QoS yang sering kali bersifat reaktif terbukti tidak lagi memadai. Oleh karena itu, industri telekomunikasi kini beralih ke solusi inovatif yang ditawarkan oleh Kecerdasan Buatan atau *Artificial Intelligence* (AI). Pemanfaatan *Machine Learning* (ML) memungkinkan sistem untuk belajar dari pengalaman masa lalu untuk memprediksi dan mengoreksi akar penyebab kegagalan, mengubah paradigma manajemen jaringan dari reaktif menjadi proaktif [3]. Salah satu tantangan utama yang dihadapi oleh operator adalah prediksi *downlink throughput*, yang dianggap sebagai salah satu metrik paling memadai karena hubungannya yang erat dengan kepuasan pengguna [1].

Berbagai penelitian telah menerapkan beragam algoritma ML untuk tugas prediksi *throughput*. Di antara berbagai metode, model berbasis pohon keputusan (*tree-based models*), khususnya *Random Forest* (RF), secara konsisten menunjukkan kinerja yang sangat unggul. Sebuah studi oleh Fernandez (2024) yang membandingkan *Linear Regression* dengan *Random Forest* menemukan bahwa RF secara signifikan lebih superior dalam

memprediksi *downlink throughput*. Keunggulan ini juga dikonfirmasi oleh Al-Thaedan dkk. (2024) yang berhasil mencapai akurasi prediksi (R^2) sangat tinggi, berkisar antara 89% hingga 96% menggunakan RF di tiga operator seluler berbeda [5]. Keberhasilan model prediksi sangat bergantung pada pemilihan fitur input yang relevan, di mana mayoritas penelitian menggunakan parameter kualitas sinyal radio yang diperoleh dari pengukuran lapangan (*drive test*) sebagai standar industri [6]. Studi analisis menegaskan bahwa parameter seperti *Reference Signal Received Power* (RSRP), *Reference Signal Received Quality* (RSRQ), dan *Signal-to-Interference-plus-Noise Ratio* (SINR) memiliki pengaruh signifikan terhadap kinerja jaringan.

Tinjauan pustaka menegaskan bahwa *Random Forest* merupakan metode yang akurat untuk memprediksi *downlink throughput* pada jaringan 4G LTE berdasarkan data *drive test*. Namun, keragaman hasil penelitian sebelumnya menunjukkan bahwa pemilihan algoritma dan konfigurasi sangat bergantung pada dataset dan tujuan spesifik, sehingga proses penyempurnaan lebih lanjut diperlukan untuk mencapai performa model yang optimal. Sukemi dkk. (2023) menunjukkan bahwa penerapan *hyperparameter tuning* pada model RF berhasil meningkatkan akurasi secara signifikan, yang menggarisbawahi pentingnya optimisasi model untuk mendapatkan hasil prediksi terbaik [7]. Berdasarkan celah tersebut, penelitian ini berfokus pada pengembangan model prediksi *downlink throughput* pada jaringan 4G LTE dengan menggunakan metode *Random Forest Regression* yang teroptimasi. Tujuan utama dari penelitian ini adalah: 1) Membangun model *Random Forest Regression* untuk memprediksi *downlink throughput* berdasarkan data *drive test*; 2) Menganalisis pengaruh berbagai metode pra-pemrosesan data, optimisasi hiperparameter $n_estimators$, dan rasio pembagian data untuk menemukan konfigurasi model yang paling akurat; dan 3) Mengevaluasi dan mengidentifikasi model terbaik dengan tingkat kesalahan (*Root Mean Square Error* - RMSE) terendah.

2. TINJAUAN PUSTAKA

BAB ini menguraikan konsep-konsep kunci yang digunakan dalam penelitian, mulai dari parameter jaringan 4G LTE, metode pengumpulan data, hingga algoritma *machine learning* dan metrik evaluasi yang diterapkan.

2.1. Parameter Kualitas Jaringan dan Pengukurannya

Kinerja jaringan 4G LTE dapat diukur melalui beberapa *Key Performance Indicator* (KPI) yang merepresentasikan kualitas koneksi dari perspektif pengguna. Dalam penelitian ini, parameter-parameter tersebut dikumpulkan melalui metode *drive test*, yaitu sebuah proses standar untuk pengumpulan data kinerja jaringan seluler secara langsung dari lapangan dalam kondisi nyata[8]. Parameter utama yang digunakan adalah sebagai berikut:

- a. *Reference Signal Received Power* (RSRP) adalah representasi kekuatan sinyal yang diterima dari sinyal referensi pada sel penyaji (*-serving cell*)[9]. RSRP diukur dalam satuan dBm dan menjadi parameter vital untuk prosedur mobilitas jaringan seperti *handover*[10]. Klasifikasi kualitas sinyal berdasarkan RSRP disajikan pada Tabel 1.

Tabel 1. Klasifikasi Kualitas Sinyal Berdasarkan Nilai RSRP

Kategori	RSRP (dBm)	Kualitas
<i>Excellent</i>	> -80	Sangat Baik
<i>Good</i>	-80 hingga -90	Baik
<i>Marginal</i>	-90 hingga -100	Cukup
<i>Poor</i>	-100	Buruk

Sumber: Shakir dkk. (2023)

- b. *Reference Signal Received Quality* (RSRQ) merepresentasikan kualitas sinyal referensi yang diterima dengan mempertimbangkan derau (*noise*) dan interferensi dari sel-sel lain di sekitarnya [12]. RSRQ dihitung sebagai rasio antara RSRP dengan *Received Signal Strength Indicator* (RSSI) [13]. Nilai yang semakin

mendekati nol menunjukkan kualitas sinyal yang lebih baik [14]. Rumus dan klasifikasi RSRQ disajikan di bawah ini.

$$RSRQ = N \left(\frac{RSRP}{RSSI} \right) \quad (1)$$

Tabel 2. Klasifikasi Kualitas Sinyal Berdasarkan Nilai RSRQ

Kategori	Rentang (dBm)
Sangat Baik	-10 s.d -3
Baik	-12 s.d -10
Cukup	-14 s.d -12
Buruk	-17 s.d -14
Sangat Buruk	-20 s.d -17

Sumber: Budiman & Hairah (2021)

- c. *Signal-to-Interference-plus-Noise Ratio* (SINR) adalah rasio antara daya sinyal yang diinginkan dengan jumlah daya dari sinyal pengganggu (*interference*) dan derau (*noise*) [13]. Parameter ini menjadi tolok ukur utama kejernihan sinyal dan berdampak langsung pada kecepatan transfer data (*throughput*) [15].

$$SINR(dB) = 10 \log_{10} \left(\frac{S}{I + N} \right) \quad (2)$$

Tabel 3. Klasifikasi Kualitas Sinyal Berdasarkan Nilai SINR

Kategori	Rentang Nilai (dB)	Kualitas
Excellent	≥20	Sangat Baik
Good	10 hingga 20	Baik
Fair	0 hingga 10	Cukup
Poor	<0	Buruk

Sumber: AL-Nasrawi dkk. (2024)

- d. *Downlink Throughput* adalah kecepatan transfer data aktual dari menara seluler (eNodeB) ke perangkat pengguna (UE), yang umumnya dinyatakan dalam bit per detik (bps) [16]. Parameter ini secara langsung mencerminkan pengalaman pengguna, sehingga menjadi variabel target yang akan diprediksi dalam penelitian ini.

Gambar 2. Tahapan Penelitian

3.2. Teknik Pengumpulan Data

Data primer dalam penelitian ini dikumpulkan melalui metode *drive test* di lapangan. *Drive test* adalah proses standar untuk merekam parameter kinerja jaringan seluler dalam kondisi nyata. Pengambilan data dilakukan di area Karangwangi, Kecamatan Cidaun, Kabupaten Cianjur, Jawa Barat. Proses ini menggunakan *smartphone* Android yang dilengkapi aplikasi G-NetTrack Pro untuk merekam parameter-parameter berikut secara simultan:

- a. Parameter Kualitas Sinyal: RSRP, RSRQ, dan SINR.
- b. Variabel Target: *Downlink Throughput* (dicatat sebagai DL_bitrate).
- c. Data Lokasi dan Waktu: *Longitude*, *Latitude*, dan *Timestamp* untuk setiap titik pengukuran.

Seluruh data yang terekam disimpan dalam format .csv untuk diolah lebih lanjut. Dataset awal terdiri dari 471 baris data dan 7 kolom fitur. Sampel data mentah yang digunakan disajikan pada Tabel 5.

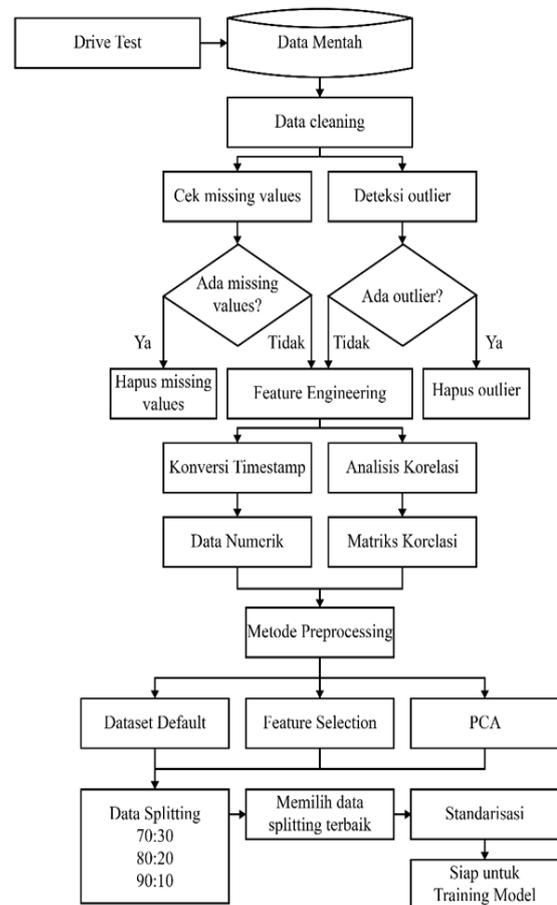
Tabel 5. Sampel Data Mentah Hasil *Drive Test*

Timestamp	Latitude	Longitude	RSRP	RSRQ	SINR	Downlink k (kbps)
2024-06-14 13:04:04	- 7.465056	107.391059 1	-92.11	-16.04	6	55113.68
	83					
2024-06-14 13:04:05	- 7.465011	107.391059 1	-94.86	-18.11	4.75	55113.68
	92					
2024-06-14 13:04:06	-7.464967	107.391059 1	-94.84	-17.48	5	52352.43
2024-06-14 13:04:07	- 7.464922	107.391059 1	-95.71	-16.26	4.25	52352.43
	09					

3.3. Pra-pemrosesan dan Analisis Data

Pra-pemrosesan adalah tahap krusial untuk mengubah data mentah menjadi format yang bersih, konsisten, dan optimal untuk pemodelan. Proses ini, yang diilustrasikan pada

Gambar 3, meliputi beberapa langkah utama.



Gambar 3. Proses Pra-pemrosesan Data

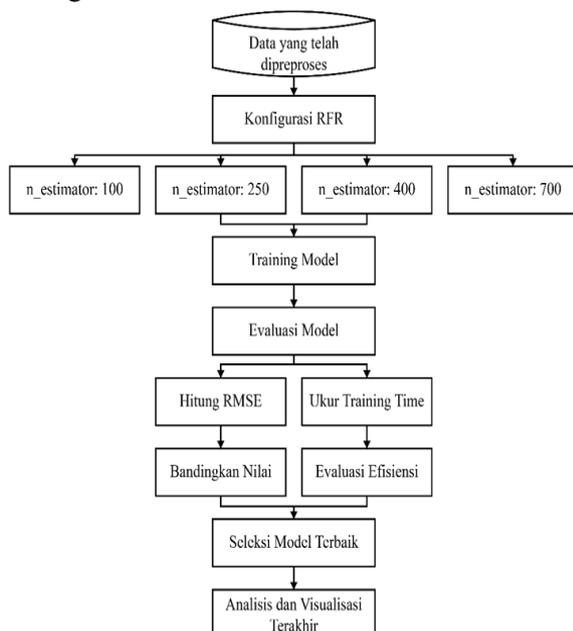
- a. Pembersihan Data: Data diperiksa untuk memastikan tidak ada nilai yang hilang (*missing values*). Selanjutnya, deteksi data pencilan (*outlier*) dilakukan menggunakan metode *Interquartile Range* (IQR) pada semua fitur numerik. Berdasarkan analisis, tidak ditemukan adanya *outlier* yang signifikan dalam dataset.
- b. Rekayasa dan Analisis Fitur: Kolom *Timestamp* yang awalnya bertipe data objek (teks) dikonversi menjadi format numerik *POSIX time* agar dapat diproses oleh model. Analisis korelasi menggunakan koefisien Pearson dilakukan untuk memahami hubungan linear antar variabel.
- c. Perbandingan Skenario Pra-pemrosesan: Untuk menemukan pendekatan terbaik, penelitian ini membandingkan tiga skenario pra-pemrosesan yang berbeda:

- Dataset Default: Menggunakan seluruh fitur yang tersedia (Timestamp, Longitude, Latitude, RSRP, RSRQ, SINR).
- Dataset *Feature Selection*: Hanya menggunakan tiga parameter kualitas sinyal (RSRP, RSRQ, SINR) dengan menghilangkan fitur lokasi.
- Dataset PCA (*Principal Component Analysis*): Mereduksi tiga fitur sinyal (RSRP, RSRQ, SINR) menjadi satu komponen utama tunggal bernama 'performance' yang mampu menangkap 85,1% dari total informasi gabungan.

Pembagian dan Standardisasi Data: Dataset dibagi menjadi data latih dan data uji dengan tiga variasi proporsi (70:30, 80:20, dan 90:10) untuk menemukan rasio paling optimal. Setelah itu, fitur-fitur numerik distandardisasi menggunakan StandardScaler dari library scikit-learn. Proses fit pada scaler hanya dilakukan pada data latih untuk mencegah kebocoran data (*data leakage*).

3.4. Pemodelan Random Forest Regression

Tahap pemodelan mengimplementasikan algoritma *Random Forest Regression* (RFR) untuk memprediksi *downlink throughput*. Algoritma ini dipilih karena kemampuannya menangani hubungan non-linear yang kompleks, akurasi tinggi, dan ketahanannya terhadap *overfitting*. Alur kerja pemodelan dirangkum dalam Gambar 4.



Gambar 4. Proses Pemodelan *Random Forest*

Pelatihan model dilakukan dengan optimisasi hiperparameter. Penelitian ini secara spesifik menguji pengaruh jumlah pohon (*n_estimators*) dengan tiga konfigurasi: 250, 400, dan 700. Hiperparameter lain seperti *max_depth* diatur ke 16 dan *random_state* ke 55 untuk memastikan hasil yang dapat direproduksi. Seluruh implementasi dilakukan menggunakan Google Colab dengan library scikit-learn, pandas, dan lainnya.

3.5. Teknik Evaluasi Model

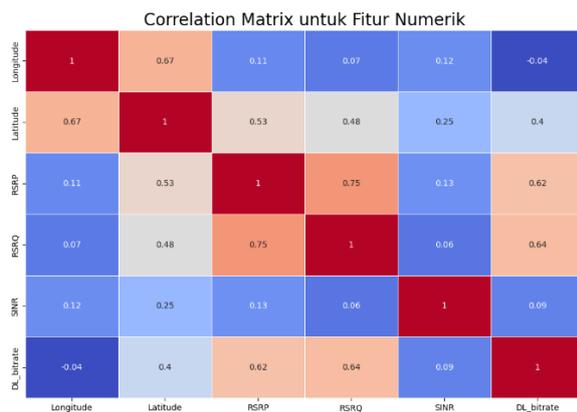
Kinerja model dievaluasi secara kuantitatif menggunakan metrik *Root Mean Square Error* (RMSE) untuk mengukur tingkat kesalahan prediksi. Nilai RMSE yang lebih rendah menunjukkan performa model yang lebih baik. Selain RMSE, durasi komputasi juga dicatat untuk menilai efisiensi model. Terakhir, untuk memastikan keandalan model, dilakukan uji stabilitas dengan menjalankan proses latih-uji sebanyak 1000 kali (iterasi).

4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari serangkaian analisis dan pemodelan yang telah dilakukan serta pembahasan mendalam mengenai temuan tersebut. Pembahasan diuraikan secara sistematis untuk menjawab rumusan masalah penelitian, mulai dari analisis eksplorasi data, optimisasi konfigurasi model, hingga evaluasi dan validasi model terbaik yang dihasilkan.

4.1. Analisis Data Eksplorasi dan Korelasi Fitur

Tahap awal penelitian berfokus pada analisis eksplorasi untuk memahami karakteristik dataset. Analisis korelasi menggunakan koefisien Pearson dilakukan untuk mengukur hubungan linear antara fitur-fitur input dengan variabel target (*downlink throughput*). Visualisasi hasil korelasi disajikan dalam bentuk *heatmap* pada Gambar 5.



Gambar 5. Heatmap Korelasi Antar Fitur Numerik

Berdasarkan *heatmap*, ditemukan bahwa fitur RSRQ (0.64) dan RSRP (0.62) memiliki korelasi positif yang paling kuat dengan *downlink throughput*. Temuan ini sejalan dengan penelitian-penelitian sebelumnya yang juga mengidentifikasi parameter kualitas dan kekuatan sinyal sebagai prediktor utama. Hal ini mengonfirmasi bahwa seiring membaiknya kualitas dan kekuatan sinyal yang diterima oleh perangkat, kecepatan unduh cenderung meningkat secara signifikan. Selain itu, terdapat korelasi positif yang sangat kuat antara RSRP dan RSRQ (0.75), yang wajar terjadi karena keduanya merupakan parameter yang mengukur aspek sinyal referensi.

4.2. Optimisasi Konfigurasi Model Prediksi

Untuk menjawab tujuan penelitian dalam menemukan konfigurasi model yang paling akurat, serangkaian eksperimen dilakukan untuk mengoptimalkan tiga aspek kunci: proporsi pembagian data, jumlah *trees* pada *Random Forest*, dan metode pra-pemrosesan data.

a. Penentuan Proporsi Data Latih dan Uji

Pengujian dilakukan terhadap tiga rasio pembagian data (70:30, 80:20, dan 90:10) untuk menemukan keseimbangan terbaik antara proses belajar model dan evaluasi yang objektif. Hasil evaluasi berdasarkan nilai RMSE pada data uji disajikan pada Tabel 6.

Tabel 6. Perbandingan RMSE Berdasarkan Proporsi Pembagian Data

Proporsi	RMSE Test
70:30	9201.575
80:20	11676.41
90:10	24975.08

Hasil pada Tabel 6 menunjukkan bahwa proporsi **70:30** menghasilkan nilai RMSE terendah secara signifikan (9201.575). Hal ini mengindikasikan bahwa rasio ini memberikan keseimbangan terbaik, di mana model mendapatkan data latih yang cukup untuk belajar tanpa mengalami *overfitting*, sekaligus menyediakan data uji yang representatif untuk evaluasi yang valid. Oleh karena itu, rasio 70:30 ditetapkan sebagai konfigurasi terbaik untuk semua pengujian selanjutnya.

b. Analisis Pengaruh Jumlah *Trees* (*n_estimators*)

Selanjutnya, dilakukan optimisasi hiperparameter *n_estimators* dengan menguji tiga konfigurasi (250, 400, dan 700) untuk menemukan jumlah pohon yang paling akurat.

Tabel 7. Perbandingan Kinerja Berdasarkan Jumlah *Trees*

Jumlah Trees	RMSE Test
trees=250	9219.719
trees=400	9232.858
trees=700	9201.575

Dari Tabel 7, konfigurasi dengan 700 *trees* terbukti memberikan akurasi prediksi tertinggi dengan nilai RMSE terendah, yaitu 9201.575. Meskipun penambahan jumlah pohon dapat meningkatkan beban komputasi, pencapaian akurasi tertinggi menjadi prioritas utama penelitian ini. Dengan demikian, *n_estimators*=700 dipilih sebagai konfigurasi optimal.

c. Analisis Metode Pra-pemrosesan Data

Tahap optimisasi terakhir adalah membandingkan tiga skenario pra-pemrosesan data untuk menjawab pertanyaan penelitian mengenai pengaruhnya terhadap akurasi model. Tiga skenario yang diuji adalah Default (menggunakan semua fitur), *Feature Selection* (hanya fitur sinyal), dan PCA (reduksi dimensi fitur sinyal).

Tabel 8. Perbandingan Kinerja Berdasarkan Metode Pra-pemrosesan

	RMSE	Duration
Default	9201.57544	3851.81499

Feature Selection	13840.336	2780.04599
PCA	18258.1889	1904.1402

Hasil pada Tabel 8 menunjukkan temuan yang sangat signifikan: model yang dilatih menggunakan dataset Default (memanfaatkan semua fitur) secara mutlak memberikan performa terbaik dengan RMSE terendah. Pendekatan yang bertujuan menyederhanakan model, seperti *Feature Selection* dan PCA, justru menghasilkan tingkat kesalahan yang jauh lebih tinggi. Temuan ini memberikan pemahaman baru yang memajukan diskusi dari tinjauan pustaka: ternyata, fitur lokasi (Longitude dan Latitude), meskipun bukan parameter sinyal, membawa informasi kontekstual spasial yang krusial bagi model untuk membuat prediksi akurat. Menghilangkannya, seperti pada *Feature Selection*, atau mentransformasikannya seperti pada PCA, terbukti menyebabkan hilangnya informasi penting yang menurunkan akurasi model secara drastis.

4.3. Evaluasi dan Validasi Model Terbaik

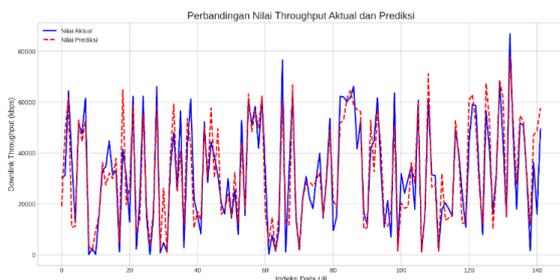
Berdasarkan serangkaian optimisasi di atas, konfigurasi model terbaik diidentifikasi sebagai berikut:

- Metode Pra-pemrosesan: Dataset Default (menggunakan semua fitur).
- Proporsi Data: 70% Latih, 30% Uji.
- Hiperparameter $n_estimators$: 700.

Model terbaik ini kemudian dievaluasi lebih lanjut untuk memvalidasi performa dan stabilitasnya.

a. Analisis Visual Prediksi vs. Aktual

Grafik perbandingan antara nilai *throughput* aktual (garis biru) dengan nilai hasil prediksi (garis merah) disajikan pada Gambar 6.

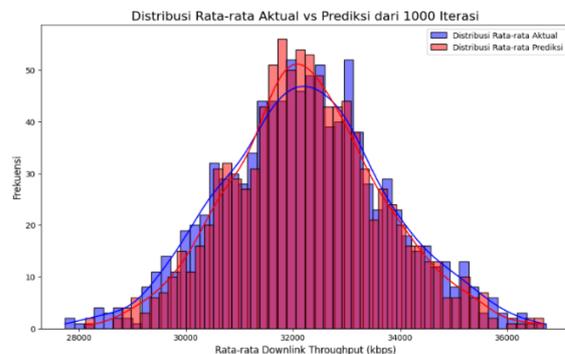


Gambar 6. Grafik Perbandingan Nilai Aktual dan Prediksi *Throughput*

Dari visualisasi, terlihat bahwa model prediksi secara umum mampu mengikuti tren utama dari data aktual, menandakan model berhasil mempelajari hubungan mendasar antara fitur input dan *throughput*. Namun, model cenderung menghasilkan garis yang lebih halus dan kurang mampu mereplikasi fluktuasi tajam serta nilai ekstrem (puncak dan lembah) pada data aktual. Fenomena ini wajar untuk model regresi berbasis ansambel seperti *Random Forest* yang mengutamakan stabilitas.

b. Validasi Stabilitas Model dengan Uji Iteratif

Untuk memastikan keandalan model tidak hanya bergantung pada satu kali pembagian data secara kebetulan, dilakukan uji stabilitas dengan menjalankan proses latih-uji sebanyak 1000 kali iterasi. Pada setiap iterasi, data dibagi secara acak, kemudian model dilatih dan dievaluasi untuk memastikan kinerjanya konsisten secara statistik. Hasil dari 1000 iterasi tersebut dirangkum untuk melihat distribusi dari nilai rata-rata *throughput* aktual dan prediksi, yang divisualisasikan pada Gambar 7.



Gambar 7. Distribusi Rata-rata Aktual vs Prediksi dari 1000 Iterasi

Pada Gambar 7, terlihat jelas bahwa distribusi nilai rata-rata aktual (biru) dan prediksi (merah) menunjukkan tumpang tindih (*overlap*) yang sangat signifikan. Keduanya membentuk kurva distribusi normal yang hampir identik, yang mengindikasikan bahwa secara konsisten, model mampu menghasilkan prediksi dengan rata-rata yang sangat mendekati nilai rata-rata aktualnya.

Analisis kuantitatif dari pengujian ini semakin memperkuat temuan visual tersebut:

- Rata-rata Nilai Aktual: 32222.05 kbps
- Rata-rata Nilai Prediksi: 32253.18 kbps
- Selisih Rata-rata: Hanya 31.13 kbps

Perbedaan yang sangat kecil ini menjadi bukti kuat bahwa model prediksi yang dibangun tidak hanya akurat dalam satu skenario pengujian, tetapi juga sangat stabil untuk memberikan estimasi yang konsisten.

5. KESIMPULAN

Berdasarkan serangkaian analisis dan hasil pemodelan yang telah diuraikan, penelitian ini berhasil menjawab seluruh rumusan masalah yang diajukan. Pertama, model untuk memprediksi *downlink throughput* pada jaringan 4G LTE berhasil dibangun menggunakan metode *Random Forest Regression* dengan memanfaatkan data hasil *drive test*, yang mencakup parameter kualitas sinyal (RSRP, RSRQ, SINR) dan data lokasi. Kedua, terbukti bahwa metode pra-pemrosesan data dan optimisasi hiperparameter memiliki pengaruh yang sangat signifikan terhadap performa model. Konfigurasi model terbaik yang menghasilkan tingkat kesalahan prediksi terendah (RMSE) sebesar 9201.575 kbps diperoleh melalui kombinasi penggunaan seluruh fitur yang tersedia (*default dataset*), proporsi pembagian data 70:30, dan hiperparameter *n_estimators* sebanyak 700 *trees*. Validasi melalui 1000 iterasi pengujian juga mengonfirmasi stabilitas model yang sangat tinggi, dengan selisih rata-rata antara nilai aktual dan prediksi hanya sebesar 31.13 kbps. Signifikansi utama dari temuan ini adalah penegasan bahwa fitur lokasi (Longitude dan Latitude) membawa konteks spasial yang krusial bagi akurasi prediksi, di mana penghilangan fitur ini justru menurunkan performa model secara drastis. Hal ini memberikan pemahaman baru bahwa untuk prediksi *throughput* yang akurat, analisis tidak cukup hanya berfokus pada parameter teknis sinyal saja, melainkan juga harus memperhitungkan konteks geografis.

Meskipun model yang dikembangkan terbukti akurat dan stabil, teridentifikasi adanya keterbatasan dalam mereplikasi fluktuasi data yang sangat tajam. Oleh karena itu, penelitian di masa depan disarankan untuk mengeksplorasi

algoritma yang lebih unggul dalam menangani data deret waktu, seperti *Long Short-Term Memory* (LSTM). Untuk meningkatkan kemampuan generalisasi model, disarankan pula untuk menggunakan dataset yang lebih beragam dari berbagai jenis lingkungan (misalnya, perkotaan padat atau dalam ruangan). Terakhir, penelitian selanjutnya dapat diperkaya dengan menambahkan fitur-fitur kontekstual baru seperti informasi beban sel (*cell load*), klasifikasi waktu (jam sibuk), atau parameter *handover* untuk memberikan pemahaman yang lebih kaya bagi model dan berpotensi meningkatkan akurasi prediksi lebih lanjut.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada Ibu Vera Noviana Sulistyawan dan Bapak Sri Sukamta atas bimbingan, diskusi, dan arahan yang konstruktif selama pelaksanaan penelitian dan penulisan artikel ini. Dukungan akademis dari Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang juga sangat berperan penting dalam kelancaran penelitian ini. Penulis juga berterima kasih kepada pihak-pihak lain yang telah berkontribusi dan memberikan dukungan dalam penyelesaian artikel ini.

DAFTAR PUSTAKA

- [1] D. Verma, H. Saraf, dan S. H. Gupta, "Prediction of user throughput from Network Parameters of LTE Network using machine learning," *Mobile Networks and Applications*, vol. 28, no. 1, hlm. 244–253, Feb 2023, doi: 10.1007/s11036-022-01934-6.
- [2] A. H. Simanjuntak, L. Nurpulaela, B. Sulisty, R. A. Faizal, dan H. M. Louhanapessy, "Pengujian Kualitas Layanan Internet Seluler Berbasis QoS Studi Kasus di Karawang," *Jurnal Profesi Insinyur Universitas Lampung*, vol. 6, no. 1, Jun 2025, doi: 10.23960/jpi.v6n1.156.
- [3] B. Bernabe, D. D. E. Michel, C. A. Marie, dan M. T. S. Fabrice, "Comparing Machine Learning Algorithms for Improving the Maintenance of LTE Networks Based on Alarms Analysis," *Journal of Computer and Communications*, vol. 10, no. 12, hlm. 125–137, 2022, doi: 10.4236/jcc.2022.1012010.
- [4] G. A. Fernandez, "Machine Learning for Wireless Network Throughput Prediction," *Advances in Machine Learning & Artificial*

- Intelligence*, no. 1, hlm. 1–6, Jan 2024, [Daring]. Tersedia pada: https://scholarworks.utrgv.edu/mss_fac
- [5] A. Al-Thaedan *dkk.*, “A machine learning framework for predicting downlink throughput in 4G-LTE/5G cellular networks,” *International Journal of Information Technology (Singapore)*, vol. 16, no. 2, hlm. 651–657, Feb 2024, doi: 10.1007/s41870-023-01678-w.
- [6] A. L. Imoize, K. Orolu, dan A. A. A. Atayero, “Analysis of key performance indicators of a 4G LTE network based on experimental data obtained from a densely populated smart city,” *Data Brief*, vol. 29, Apr 2020, doi: 10.1016/j.dib.2020.105304.
- [7] S. Sukemi, A. Oklilas, M. Fadli, dan B. Alfaresi, “Path Loss Prediction Accuracy Based on Random Forest Algorithm in Palembang City Area,” *Jurnal Nasional Teknik Elektro*, vol. 12, no. 1, Mar 2023, doi: 10.25077/jnte.v12n1.1051.2023.
- [8] S. D. Yusuf, S. I. Isa, dan B. J. Kwaha, “Analysis of 4G/LTE Network Performance in North-Central Nigeria: A Comprehensive Drive Test Approach,” *Journal of Engineering Research and Reports*, vol. 26, no. 9, hlm. 105–122, Sep 2024, doi: 10.9734/jerr/2024/v26i91267.
- [9] M. Ayad, R. Alkanhel, K. Saoudi, M. Benziane, S. Medjedoub, dan S. S. M. Ghoneim, “Evaluation of Radio Communication Links of 4G Systems,” *Sensors*, vol. 22, no. 10, Mei 2022, doi: 10.3390/s22103923.
- [10] M. Mohamed Yunus, K. Nazihah Azahar, M. Haslinda Mohamad, dan Y. Khaidir, “Semarak International Journal of Electronic System Engineering Evaluating 4G Network Performance at UTeM Campus to Facilitate 5G Implementation in Malaysia,” *Semarak International Journal of Electronic System Engineering*, vol. 3, hlm. 15–27, 2024, doi: 10.37934/sijese.3.1.1527b.
- [11] Z. D. Shakir, J. Zec, I. Kostanic, A. Al-Thaedan, dan M. E. M. Salah, “User equipment geolocation depended on long-term evolution signal-level measurements and timing advance,” *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, hlm. 1560–1569, Apr 2023, doi: 10.11591/ijece.v13i2.pp1560-1569.
- [12] E. Budiman dan U. Hairah, “Kinerja Jaringan 4G LTE Operator Mobile di Ibukota Kalimantan Timur dimasa Pandemi Covid19,” *Jurnal_Pekommas*, hlm. 1–9, Okt 2021, doi: 10.30818/jpkm.2021.
- [13] M. AL-Nasrawi, S. V. A. D. Makki, dan A. Al-Sabbagh, “KPI analysis of 4G/5G networks,” *Przeglad Elektrotechniczny*, no. 7, hlm. 67–71, 2024, doi: 10.15199/48.2024.07.15.
- [14] F. Farida dan S. Nugraha, “Analisis Kualitas Jaringan 4G LTE dengan Metode Drive Test pada Jalur Pelayaran Tanjungpinang-Batam,” dalam *Prosiding Seminar Nasional Teknik Elektro*, 2020.
- [15] J. Abdullahi dan A. A. Abdulhamid, “Evaluating 4G Network Performance in North-Central Nigeria: A Drive Test-Based Assessment of Key Performance Indicators Mustapha Agwai I Polytechnic (IMAP), Lafia Nasarawa State Nigeria,” vol. 7, no. 3, hlm. 305–323, Mar 2025, doi: 10.51244/IJRSI.
- [16] D. Melenia, U. Usman, dan G. Satrya, “Analisis Perbandingan Throughput Open RAN 4G LTE Arah Downlink Secara Real Dan Berdasarkan 3GPP,” vol. 8, no. 6, hlm. 2716–2722, Des 2022.
- [17] H. Elsherbiny, H. M. Abbas, H. Abou-Zeid, H. S. Hassanein, dan A. Noureldin, “4G LTE Network Throughput Modelling and Prediction,” dalam *Proceedings - IEEE Global Communications Conference, GLOBECOM*, 2020. doi: 10.1109/GLOBECOM42002.2020.9322410.
- [18] H. Yuliana *dkk.*, “Hyperparameter Optimization of Random Forest for 5G Coverage Prediction,” *Buletin Pos dan Telekomunikasi*, vol. 22, no. 1, Jun 2024, doi: 10.17933/bpostel.v22i1.390.
- [19] G. Biau dan E. Scornet, “A random forest guided tour,” *Test*, vol. 25, no. 2, hlm. 197–227, Jun 2016, doi: 10.1007/s11749-016-0481-7.